

372.05
CO
V. 1:1

The Cornbelt Education Review A Graduate Student Journal

The Library of the

AUG 30 1982

*University of Illinois
at Urbana-Champaign*

Department of Elementary & Early Childhood Education
College of Education
University of Illinois
Urbana, Illinois

Volume 1

Number 1

Editorial Guidelines

With this issue, the Cornbelt Education Review is enjoying a modest beginning, with its small selection of articles and limited circulation. As a graduate journal, it is intended to provide an opportunity for professional sharing of the research efforts and concerns of beginning researchers. We are endeavoring to encourage the preparation of publications, the flexing of editorial skills, and the use of research methodologies which may not be completely developed or even widely accepted within the academic community as a whole, naturalistic or experimental research, book reviews, evaluations. The Cornbelt Education Review offers graduate students space to explore research methodologies alternative to those perhaps required in the dissertation. More importantly, these attempts to translate theory into practice help underscore some of the prominent educational issues which are currently under discussion in the College and beyond.

Editorial Guidelines for the journal suggest the inclusion of or surveys by graduate students from Elementary and Early Childhood Education. Articles of approximately 2000-4000 words in APA style should be submitted to the Editorial Board, 330 Education Building for refereed selection.

We wish to acknowledge and express appreciation for the efforts of Dean J. Myron Atkin, Department Chairman Harold Lerch and the journal Advisor Bernard Spodek, who were instrumental in initiating its conceptualization and supported its development.

The Editorial Staff

Judith Pickle
Elizabeth Rosen
Patricia Scheyer

The Cornbelt Education Review

Volume 1 Issue 1
June 1977

Table of Contents

Letter From The Dean Dean Myron Atkin	XII
Protocols - A Review Synthesis Nancy Blair	1
Reflection In Student Teaching James L. Hoot	10
Test Results Revisited Patricia Scheyer	20
Conversations With Myself or "What's A Nice Girl Like You Doing in Evaluation" Carol Hodges Wardrop	32

The Cornbelt Education Review is sponsored by the Department of Elementary and Early Childhood Education at the University of Illinois.



Digitized by the Internet Archive
in 2011 with funding from
University of Illinois Urbana-Champaign

Congratulations to the Department of Elementary and Early Childhood Education for launching the Corn Belt Education Review. Congratulations are in order not so much because the world needs another education journal but because the voices of a strong and influential Department just might have an impact at a time when its field is beset by confusion and uncertainty. The Department of Elementary and Early Childhood Education at the University of Illinois at Urbana-Champaign has a distinguished history; it is staffed currently by an extraordinarily strong faculty; it continues to attract some of the most able graduate students in the College. Perhaps this group can speak to issues associated with accountability in education, with the state of our teacher education programs across the country, with the challenges of continuing education for classroom teachers, with racial conflict and bilingualism in the schools, with curriculum questions, and with the highly-charged issue of selection for the teaching profession.

In most departments of elementary and early childhood education, the faculty and students are preoccupied with declining enrollments and budgets. Fortunately, fifteen years ago, this Department decided to limit undergraduate teacher education numbers at the University of Illinois at Urbana-Champaign and focus on graduate-level training, research, and development. Partly as a result, the Department is probably in as strong a position programmatically as any, anywhere.

For all these reasons the initiation of the Corn Belt Education Review is an event that offers promise for according thoughtful people the opportunity to reflect on current problems from both a privileged and a distinguished platform. I look forward to seeing the results.

J. Myron Atkin, Dean
April 1977

PROTOCOLS--A REVIEW SYNTHESIS

Nancy Blair
Elementary Education

Introduction

B. O. Smith's Teachers for the Real World was the catalyst for the protocol materials development movement. Smith voiced a concern for the failure of teacher training institutions to prepare teachers to analyze new situations against a background of relevant theory. Smith states that teachers, "typically...base their interpretations of behavior on intuition and common sense...If the teacher is incapable of understanding classroom situations the actions he takes will often increase his difficulties." (Smith, 1969)

Protocols are original records of events of educational significance. They are utilized to permit teachers or prospective teachers to interpret events with the aid of appropriate concepts from related fields of study. (Cruikshank, 1974) They are reproductions of student, teacher or parent behavior recorded on film, cassette or filmstrips. They exemplify educational concepts and are designed to teach the application of these concepts and to train school personnel to analyze and classify behavior.

Protocols offer several advantages over direct observation in real classrooms. In real life, behaviors are fleeting. They may be missed or they may be recalled faultily for later analysis. With the use of protocols a behavior can be captured and replayed as many times as necessary for study and analysis. Thus, protocols are more efficient for analysis since particular behaviors and related concepts can be utilized at appropriate times in the students training. (Gunderson, 1974) On the other hand, protocols lack the spontaneity and some of the "realness" of direct observation. Ideally, protocols would provide examples of behavior that would help the observer acquire the conceptual knowledge necessary to interpret pedagogical events.

Claims

Numerous claims have been made as to the value of protocol materials. Orlosky's support for the use of protocols is based on the belief that "if behaviors could be captured on film, students in training could see the behavior exemplifying a given concept and be better prepared to

recognize and classify the behavior." (Orlosky, 1974) Gunderson asserts that certain cognitions come from the exposure to protocols. She feels they teach concepts and principles used in interpreting behavior, and that they teach knowledge about knowledge as well as self-understanding. (Gunderson, 1974) Gee and Berliner have broken down the acquisition of concepts to interpret behavior into three stages: 1) functional knowledge of psychology, philosophy and social concepts relevant to teaching, 2) ability to interpret behavioral situations in terms of these concepts and 3) the ability to formulate plans for teaching based on the first two skills. In other words protocols give teachers a basis for decision making in the classroom. (Gee and Berliner, 1971)

Smith fears that there is a danger of reducing teacher training to the development of teaching skills alone. He claims that protocols can bridge the gap between theory and practice. He emphasizes interpretive and replicative uses of theoretical knowledge and the use of protocols to develop a teacher's conceptual system--using the theoretical knowledge in real situations. He states:

By being involved repeatedly in the process of analyzing and interpreting them (behavioral situations), the prospective teacher will learn to interpret quickly and thoroughly the events and episodes that happen as he teaches.

(Smith, 1969)

Beyond this, the student will also become aware of the great variety of ways of handling each teaching task as well as the different ways teaching behaviors and situations may be interpreted. Smith further claims that through the use of protocols, teachers will acquire an increased interest in theory and recognize its usefulness.

Selection of Concepts

Developers of protocols have tried to identify teaching concepts and then depict these concepts in protocol format. There are several problems attached to this approach to concept selection. First, the concepts in the foundational areas of education are weak due to our lack of knowledge about program outcomes. (Hudgins, 1972) Second, if protocols are to be selected to reflect concepts with generality, utility, and interpretive possibilities, we must first carefully analyze the concepts identified in depth. The dimensions and standards of such an analytic process are not widely understood in psychology, sociology and

anthropology. (Gleissman, 1972) Furthermore, teacher education is a difficult substantive area due to the lack of a match between concepts and the exemplary behaviors. In spite of these problems it was believed that delay in production of protocol materials pending definitive statements concerning concept selection would be counterproductive.

Before production began the concepts chosen had to be clearly and precisely defined. The Leadership Training Institute on Protocol Materials recommended that concepts be expressed in one of four possible forms; classification, equivalent expression, open context and conditional. Classification calls for the association of a concept with a category and discriminating criteria distinguishing one concept from another. In the equivalent expression form a concept is explained by a statement equivalent to word or words used to name that concept. Open context is used for concepts that are not clearly definable, such as happiness. Conditional concepts are those that can be explained only in terms of a context or conditions surrounding the event. (Smith and Orlosky, 1973)

Initially, all of the developers listed the concepts that they planned to produce in protocol form. As work progressed the initial choices were generally maintained but with greater emphasis on clarification and supporting information. The developers eventually identified concepts that are part of a system rather than isolated from each other.

Building the Concept

After a concept has been identified, it must be analyzed and evaluated. The following steps in evaluating a concept have been used.

1. Is it representable?
2. Is it significant? (high utility, critical to teacher education, and shows empirical substantiation)
3. What are the learning objectives related to this concept?
4. Who is the anticipated user?
5. How will it be field tested? (Gunderson, 1974)

In analyzing a concept, one must identify the range and types of behaviors to be captured to portray the concept fairly. It is necessary to specify the dimensions and values of the concept in order to insure comprehensive representation. To clearly define a concept one must consider complexity of examples, sequencing, range of discriminations the learner makes, definitions and descriptions. (Hudgins, 1972)

The amount of literature on concept acquisition is an indication of the complexity of this field. Jerome Bruner, in A Study of Thinking, refers to concept attainment as the "process of finding predictive defining attributes that distinguish exemplars from nonexemplars of the class one seeks to discriminate." He suggests that examples and counter examples given to students optimize the acquisition of any given concept regardless of the type of learning strategy employed by the student. (Bruner, 1956)

While Gagne agrees that simpler concepts can be learned through the contrast of positive and negative examples, he further suggests that more complex concepts (i.e., work, uncle) which are combinations of simpler concepts are concepts by definition. They are learned via verbal statements that provide cues to recall of component concepts and their correct ordering. Gagne asserts that there is much to be done in conducting research on the acquisition of these two types of concepts, and that in order for the research to be meaningful we must distinguish between simple and more complex types of concepts. (Gagne, 1966)

Hunt defines concept learning in terms of learning a rule that would apply to any appropriate stimulus that fits the given concept and this rule would be uniquely classifiable. (Hunt, 1962) This differs from Piaget's definition of a concept as an explanatory rule that need not be a classification rule. (Piaget, 1957) Thus, the issue is further complicated by the lack of agreement on terminology. Ausubel believes that building a clear, stable and unambiguous meaning of the concept in the student's cognitive structure is essential. (Ausubel, 1968)

Hudgins has tried to incorporate some of these theories on concept acquisition for use with protocol materials. His model calls for concept presentation that increases stimulus complexity.

<u>Stage I</u>	<u>Stage II</u>
Printed definition	Example on film
Written examples	Low noise
	Cues, labels
<u>Stage III</u>	<u>Stage IV</u>
Positive and negative examples	"Slice of life" (less editing)
Increased noise	
Fade cues, labels	(Hudgins, 1972)

Another issue that arises is whether or not teacher-learning episodes should be staged for filming. This has caused concern among developers but it is an empirical question that has not yet been answered. Certainly finances and time constraints support the staging of situations. (Cruikshank, 1972)

The Use of the Materials

There has been some disagreement between the heuristic and didactic use of protocols. Should students watch the protocol material and arrive at the concept inductively or would direct guidance be more efficient? Neither research nor theory lends adequate support to this distinction. (Smith, Orlosky, 1973) Perhaps the main consideration would be whether or not the concepts have been presented in such a way to increase understanding. One such avenue for increased understanding of a concept has been suggested by Innerd and O'Gorman.

1. Information--relevant theory presented as hypothesis to be tested.
2. Protocol Material--video of real behavior shown.
3. Analysis--viewer discusses behavior in relation to theory.
4. Discussion--does the theory hold up or is it denied? (Innerd, O'Gorman, 1970)

Evaluation

There are several types of evidence that have been compiled to demonstrate the effectiveness of protocol materials. Most of the evidence is reported in terms of concept acquisition and reactions to the materials by

trainees and their instructor. Most of this evidence was collected during field tests of the products rather than large-scale evaluation efforts.

Utah State University, Michigan State University, Far West Laboratory, Indiana University and Southern Illinois University at Edwardsville have reported superiority in concept growth in groups using protocol materials as compared to groups using traditional lecture-discussion and printed materials. Studies at Michigan State produce evidence not only of concept acquisition but also on its transfer to simulated teaching situations. (Cooper, 1975) It has not been demonstrated that this transfer occurred in a real classroom situation.

Instruments have been designed by Protocol Projects Materials directors to measure attitudes of students to protocol materials. These instruments gather reactions to technical quality and usefulness of content for improving teaching. These result in value judgments which for the most part are favorable to protocol products. Indiana University collected information on the reactions of instructors. Six instructors from several institutions gave positive responses to the materials' quality, appropriateness of content and usefulness in promoting intended concepts. (Cooper, 1975)

Kleuker measured the effectiveness of protocol materials and training materials (methodological and skill training) together and separately as related to concept acquisition. She, too, found that students using protocol materials were superior in concept acquisition to a control group. She also found that students who received both protocol and skill training performed better in concept recognition than students who just received protocol training. (Kleuker, 1974) One must ask whether the added exposure to the learning situation accounted for this increase in performance rather than the combination of the two approaches.

The research conducted thus far is limited and not directed at the many claims advanced in support of the use of protocols. There are many questions left unanswered. Are students trained via protocols better able to formulate plans for a learning situation based on these concepts? Do these students have increased self-understanding? Does it effect their performance as a teacher? Are these students able to quickly and thoroughly interpret the complex behaviors that occur in the classroom? As Smith claims, have the students developed an increased interest in theory due to their interaction with protocols? Have the students

developed an increased recognition of the usefulness of theory? Have there been any positive changes in pupil behavior that would support the use of protocol materials? At this stage, the evaluation efforts are not congruent with the promises made throughout the development process. Enough protocols have been developed to provide the necessary materials for large scale evaluation that addresses some of the questions raised here.

One would agree with Smith that teacher educators must go beyond concerns for technique and skill acquisition. We need to stimulate the intellect of the prospective teacher and instill in him the dispositions to be analytical, synthetical, reflective and curious. The protocol movement has been a fascinating attempt to develop materials to do just that and future investigations should determine its true merit. At the very least, encouragement is needed for careful evaluation of these materials that goes beyond concept acquisition and student reactions.

Review of the Literature

Ausubel, David, Educational psychology: a cognitive view.
Holt, Rinehart and Winston, New York, 1968.

Briggs, Leslie J., A catalogue of protocol materials in teacher education. Tallahassie, Florida, State Department of Education, 1971.

Bruner, Jerome S., Goodnow, J. J., and Austin, G. A., A Study of thinking, Wiley, New York, 1956.

Cruikshank, David R., The developing notion of protocol materials. Journal of Teacher Education, 1972, 23, 281-285.

Cruikshank, David R., The protocol movement: an exemplar of efforts to wed theory and practice. Journal of Teacher Education, 1974, 25, 300-311.

Gee, James, and Berliner, Davis, Protocols: a new dimension in teacher education. Bureau of Personnel Development, USOE, 1971.

Gliessman, David, An introduction to protocol and training materials. Acquiring Teacher Competencies, Bloomington, Indiana, National Center for the Development of Training Materials in Teacher Education, 1972.

Gunderson, Doris V., Protocol materials. National Center for the Improvement of Educational Systems, USOE, 1974.

Gagne, R. M., The learning of principles. Analysis of concept learning, Klausmeier and Harris, editors, Academic Press, New York, 1966.

Hudgins, Bryce B., Concepts in the pedagogical domain. Journal of Teacher Education, 1974, 25, 330-337.

Hudgins, Bryce, B., The portrayal of concepts: an issue in the development of protocol materials. Acquiring Teacher Competencies, Bloomington, Indiana, National Center for the Development of Training Materials in Teacher Education, 1972.

Hunt, E. B. Concept Learning, John Wiley and Sons, New York, 1962.

Kleuker, Joy, Effects of protocol and training materials,
Acquiring Teacher Competencies, Bloomington, Indiana,
National Center for the Development of Training Materials
in Teacher Education, 1974.

Orlosky, D. E., The protocol materials program. Journal of
Teacher Education, 1974, 25, 291-299.

Piaget, J., Logic and psychology, Basic Books, New York, 1957.

Smith, B. O. Orlosky, D. E., The development of protocol
materials, Acquiring Teacher Competencies, Bloomington,
Indiana, National Center for the Development of Train-
ing Materials in Teacher Education, 1973.

Smith, B. O., et al., Strategies of teaching, Bureau of
Educational Research, University of Illinois, 1967.

Smith, B. O., Cohen, Saul B., Pearl, A., Teachers for the
real world, Washington, D. C., American Association
of Colleges for Teacher Education. 1969.

REFLECTION IN STUDENT TEACHING

James L. Hoot
Early Childhood Education

In his book, Life in Classrooms, Jackson described the classroom as a "proverbial beehive of activity" (Jackson, 1968). Within this intense activity, educational decisions are made by teachers. At issue in the present study is how such decisions are made, i.e., are they made in a haphazard, thoughtless, manner--or are they the outcome of a reflective process which Dewey defined as "...active, persistent, and careful consideration of any belief or supposed form of knowledge in light of the grounds that support it..." (Dewey, 1933).

Accounts of teachers making decisions and implementing curricula with little or no thought given to assumptions underlying such implementation are not rare. From her many observations and discussions with teachers in elementary classrooms, for example, Durkin concluded that a major goal of lessons was often "filling up a day." Moreover, she stated that "While other (educational) goals were better, some were still not the kind that could easily be defended by educators" (Durkin, 1975).

In addition to concern about the assumptions underlying selection of curricular goals, recent reports have also suggested a need for teachers to become increasingly reflective about material selection. A study by the Educational Products Information Exchange Institute, for example, indicated that "...99 percent of roughly 200,000 curriculum materials on the market were not being tested prior to publication." Although these "untested" materials account for less than 1% of the national school budget (The 1974 Cost of Education Index), they often dictate a large percentage of instructional time.

The present study began with the assumption that teachers should not serve a valet function in making educational decisions. Rather, choices made by educators should be supported by a defensible belief system i.e., they should be the outcome of a reflective process. And, if a reflective disposition can be seen as a desirable teacher characteristic,

*For a complete account of this study see: James L. Hoot, "Student Teacher Reflection to Preactive, Interactive, and Reactive Teaching Concerns in Reading and Math." Unpublished Evaluation Report, University of Illinois, 1976.

it follows that institutions purporting to prepare teachers should assume greater responsibility for strengthening reflective attitudes in its students. In this view, teacher education is seen as being what Katz (1974) calls "functionally" related to what is useful in the field, i.e., "...what trainees learn in the program helps them become good at teaching."

The University of Illinois Early Childhood Teacher Education Program is presently in the process of moving toward a "liberation" model which recognizes the importance of responsible choice in educational decision-making (Spodek, 1975, Becher, 1976). With this emphasis, we have become increasingly concerned that students being recommended for certification from our program have begun to develop a propensity to reflect upon the things they do and the choices they make. Further, viewing our students as reflecting as well as assimilating agents, we would hope that the nature of our teacher education program had been instrumental in enabling students "...to make personal judgments assessing values and performance through self-reflection..." (Schickedanz and Spodek, 1975).

The present study was an outgrowth of our ongoing program evaluation process. Although limited in its scope, the study attempted to assess the degree to which a sample of our student teachers would provide evidence of reflection in planning, teaching, and evaluating sample lessons in the reading and math content areas.

Major questions guiding the present inquiry were:

1. How reflective will student teachers be in responding to interview items related to selected dimensions of the teaching process in the content areas of reading and math?
2. Will reflection scores differ between the two content areas?
3. Will reflection scores differ in content areas between the planning, teaching and evaluating aspects of teaching?
4. What factors might account for differences in reflection scores?

Procedure

In October, 1975, five University of Illinois students were randomly chosen from an eleven member section of Elementary Education 232 (Supervised Student Teaching). Students were seniors majoring in Early Childhood Education and all were completing their final year of preservice training.

While student teaching in the Champaign-Urbana school systems, subjects were asked to participate in two interview sessions which were to follow formal observations of a reading and a math lesson. The interview instrument consisted of twelve items related to planning, teaching, and evaluating the observed teaching segment.

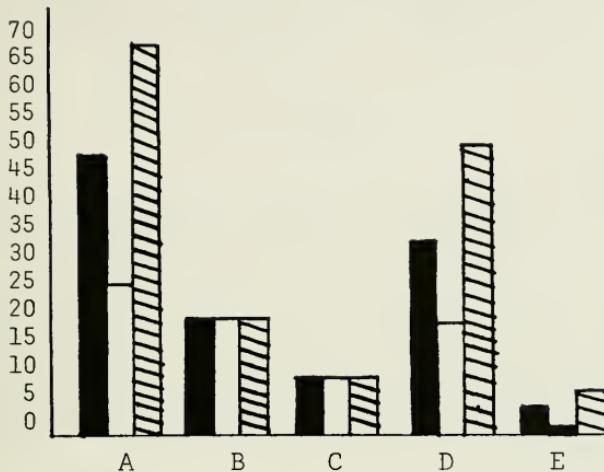
Following the interview, raters (doctoral students in Elementary/Early Childhood Education) scored responses according to the following criteria:

Reflective Response: (1 point) A response characterized by careful consideration given to the item in question. A reflective response might include such indicators as: appeal to one or more theoretic sources (e.g., learning theory, developmental theory), appropriateness for given students, societal concerns, support from internalized value judgments based on specific criteria (e.g., John and Mary understood because...).

Non-Reflective Response: (0 points) A response indicates little or no consideration was given for the item in question. Indicators might include: convenience responses such as "the materials were easy to gather," blind faith in materials, manuals and/or cooperating teachers, and global justification (e.g., "they" needed it).

Results

Figure 1 shows the percentage of responses which were rated as having provided evidence of reflection. The mean combined reflection percent was 21. Looking at combined percentages for individual subjects, students A and D were clearly rated superior.



Combined Students

Reading Figure 1

Math Percent of Responses Rated Reflective
(Combined, Reading, Math)

The mean reflection percent of the total group reading and math was 15 and 28 respectively. Analysis of individual reading/math ratings presented in Figure 1 indicates that subjects A and D were rated nearly three times more reflective in math. Subjects B and C obtained the same score for both interviews. E's score was better in math.

Figure 2 represents a comparison of reflection scores in the planning, teaching and evaluative dimensions of reading and math.

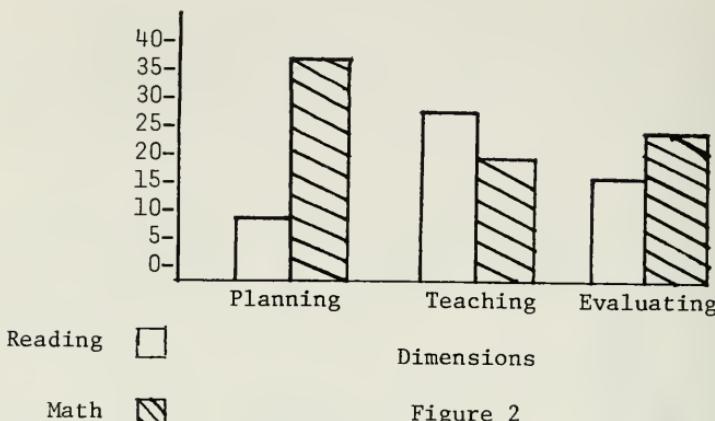


Figure 2

Reflection percent in Planning, Teaching, and Evaluative Dimensions in Reading and Math

In math, students were most reflective about planning concerns and least reflective about interactive teaching. In reading, students were most reflective about teaching aspects and least reflective about planning concerns. Chi-square analysis of independence indicated a significant difference ($p < .02$) between reading and math scores in the planning dimension.

Discussion

Before discussing the results, it must first of all be recognized that little definitive inference can be postulated from a study of this scope. Among the most notable delimitations numbered: (1) the sample was very small, (2) the interview instrument was not tested for validity, and (3) scores may have been deflated by constraints inherent in the student teaching process as described by Innaccone (1963), Jacobs (1968), Katz (1972), Fuller and Brown (1974), Cogan (1974) and others.

In answer to the research questions, results indicated that (1) subjects were rated as having responded reflectively on 21% of the items (responses of subjects A and D greatly inflated this percentage) (2) three of the five subjects scored higher in math than reading while two subjects obtained the same score in both the reading and math interviews, and (3) student responses to math interviews were rated significantly ($p < .02$) more reflective in the planning dimension than in the teaching and evaluative areas. No significant difference between reading and math scores was found in the teaching and evaluative dimensions.

Following analysis of results, an attempt was made to look at a few variables which may have accounted for differences in scores. First of all, relationships were examined between reflection scores and traditional indicators of success in formal academic course work. Looking specifically at differences between reading and math scores, a comparison was made to see if students who did well in coursework (as measured by grades) were rated more reflective in teaching the respective content than those who did less well. Table 1 shows that all students (except E) had taken formal academic coursework in both the teaching of reading (El. Ed. 336) and Math for Elementary Teachers (Math 202).

Table 1

Grades Received in Teaching of
Reading and Math

Student	A	B	C	D	E
Reading Grade	A	B	C	B	B
Math Grade	A	C	D	C	-
Reflection Rank	1	3	4	2	5

Rank-order correlation coefficients computed with Table 1 data resulted in high positive correlations between reflection rank/grade rank in reading (+.70) and reflection rank/math grade rank (+.85). With an n of five, these coefficient did not reach the critical value of .9 needed for significance at the .10 level. However, since coefficients approached the critical value, further research is needed with larger sample sizes.

Reflection score rank of each student was also paired with ratings in El. Ed. 232 (Supervised Student Teaching), El. Ed. 230 (Student Teaching Problems and Issues Seminar) and GPA. No consistent relationship between these variables was found.

In addition to examination of formal academic indicators, reflection ranks were compared with teaching variables such as: teacher-student ratios during the observed lessons, questioning styles, and ratings of cooperating teachers. Analysis of the teacher-student ratio revealed that each reading group had a ratio of 1:7 or less. In math, however, the ratio increased to 1:21 or greater for subjects who ranked 3rd, 4th, and 5th in reflection. Subjects A and D who were rated most reflective retained the 1:7 or lower ratio in the math lesson.

Questions asked by subjects during the observed lesson were rated as (1) those which elicited a single response and (2) those which elicited a range of learner responses. Data showed that subjects asked more range-response questions in math. Also, although ranked range-response questions in reading were not consistently related to reflection rank, reflection rank and range-response rank in math corresponded.

Finally, many researchers such as Perrodin (1961), Innaccone (1963), Yee (1969), Lowther (1970), Fuller and Brown (1975) and others have suggested that cooperating teachers may significantly effect attitudes and behaviors of novice pedagogues. To obtain an indication of this effect in the present study, the investigator rated each subject's cooperating teacher on three characteristics--modeling behavior, feedback provided to student teachers, and the relative amount of freedom student teachers were given to implement their ideas. Each indicator was then rated with the scale: 3 (Good), 2 (Fair), and 1 (Low). Table 2 pairs student teacher reflection rank with corresponding mean cooperating teacher rank on the above indicators.

Table 2

Student Teacher Reflection Rank and
Mean Cooperating Teacher Rank

Student Teacher	A	B	C	D	E
Student Teacher					
Reflection Rank	1	3	4	2	5
Mean Cooperating					
Teacher Rank	1	3.5	2	3.5	5

Although ranks were not perfectly correlated, student teachers with the highest and lowest reflection ranks had cooperating teachers who corresponded to those same ranks.

Summary

The present study was an attempt to obtain an indication of reflection exhibited by student teachers during the teaching process. Results indicated that: (1) subjects were rated as having responded reflectively on 21% of the total items, (2) students tended to be rated more reflective in math than reading, and (3) responses to math interviews were rated significantly ($p < .02$) more reflective than reading responses in the "planning" dimension of teaching.

Finally, if a major goal of teacher education institutions is to prepare teachers for making responsible educational decisions, the study suggested concern that program elements nurture reflective dispositions in students.

References

Becher, Rhoda. "Moving Toward a Liberation-Oriented Model in Early Childhood Teacher Education." A paper to be presented at the NAEYC Conference, Anaheim, California, 1976.

Cogan, Morris L. "Current Issues in the Education of Teachers," Kevin Ryan (ed.), Teacher Education, 74th NSSE Yearbook, Chicago: University of Chicago Press, 1975.

"Cost of Education Index 1968-74," School Management, XVIII, 1, January, 1974.

Dewey, John. How We Think., Boston: D. C. Heath, 1933.

Durkin, Dolores. "The Importance of Goals for Reading Instruction," The Reading Teacher, January, 1975.

Fuller, Francis, and Oliver Brown. "Becoming a Teacher," Kevin Ryan (ed.), Teacher Education, 74th NSSE Yearbook, Chicago: University of Chicago Press, 1975.

Innaccone, L. "Student Teaching: A Transitional Stage in the Making of a Teacher," Theory into Practice, II, 1963.

Jackson, Phillip. Life in Classrooms, New York: Holt, Rinehart and Winston, 1968.

Jacobs, E. B. "Attitude change in teacher education: an inquiry into the role of attitudes in changing teacher behavior," The Journal of Teacher Education, XIX, 1968.

Katz, L. "Issues and Problems in Teacher Education," Bernard Spodek (ed.) Teacher Education, Washington: NAEYC, 1974.

Lowther, M. A. "Successful and Unsuccessful Experiences of Student Teachers in Secondary Education." Contemporary Education, XLI, 6, May, 1970.

"More on Learner Verification," Epiegram, New York: Educational Products Information Exchange Institute, II, 9, 1974.

Perrodin, A. F. "In Support of Supervising Teacher Education Programs," Journal of Teacher Education, II, 1961.

Schickedanz, Judith and Bernard Spodek, "Conceptualizing Teacher Education Programs," A paper presented at the NAEYC Conference, Dallas, Texas, November 14, 1975.

Spodek, Bernard. "Early Childhood and Teacher Education: A Search for Consistency" Young Children, 3 March, 1975.

Yee, A. H. "Do Cooperating Teachers Influence the Attitudes of Student Teachers?" Journal of Educational Psychology, LX, 4, 1969.

TEST RESULTS REVISITED

Patricia T. Scheyer
Early Childhood Education

In many school systems, the use of standardized tests has long been the major evaluation strategy. Such assessment of student achievement, aptitude, or intelligence is usually established and regulated by state and local requirements. Though classroom teachers often administer the tests, they seldom receive results more informative than single scores for students. While some critics call for reform or abandonment of testing practices, the teachers interviewed in this study have begun developing alternative ways of evaluating children in the context of the classroom environment.

Objections to testing have been voiced in unequivocal terms by many educators. Karier (1973) scored intelligence and achievement tests as instruments of control in the corporate state, maintaining within the schools a system of sorting and classifying people into a hierarchy of social and economic classes. Zacharias (1975) claimed tests have an imprisoning and destructive effect on children when a child's worth is reduced to a single test score placed on a single line order of intelligence or achievement, particularly since there is no agreement about what test items measure.

Tests are criticized for middle-class bias, for rewarding speed and conventional thinking, for labeling children, and for measuring children's cultural and experiential backgrounds. Although there may be little agreement about what test items measure, teaching test items is one practice associated with raising students' scores. House (1973) found testing in the context of evaluation and accountability as simplistic, unworkable, contrary to empirical findings and ultimately immoral.

Test scores may be valid for discriminating among individual students within a classroom group (Stake, 1968). Their validity is low, inapplicable or unestablished for other common uses such as diagnostic indicators, comparing schools, or evaluating the competency or effectiveness of teachers or the suitability of curricula for an entire group (Stake, 1975). Tests tell the teacher little about what is happening in the program as students are learning. Hastings, Runkel, et al., (1960) found that in situations where teachers had available both test results and teacher assessments of students, they often vetoed the test results and preferred teachers' judgments instead. Hotvedt (1974) concluded that most teachers actually do not use available results.

Hoffman (1975) called for any kind of testing or evaluation which allows the student to do something himself, and where the primary concern is with the reasoning process rather than the answer. Both theoretical and applied alternatives to testing student outcome have been suggested in the literature, notably by Eisner (1969, 1974), Gallagher (1970a, 1970b), Stake (1967, 1975), Bussis, Chittendon and Amarel (1973, 1975, 1976), Perrone (1972, 1975), Carini (1975), Duckworth (1973), Wolf (1971), Scheyer and Stake (1976), Kemmis (1974), Elliott (1975, 1976), Glass (1974), Barclay (1974), and Anderson and Walberg (1974).

Program evaluators' concern has grown recently for understanding more about teaching and learning in the context of processes and interactions within the classroom environment among students, teachers, curriculum and materials. This potentially rich source of contextual data was explored in this study by asking teachers how they feel about the policy and practice of testing and the use of test results. The teachers' comments support many of the claims and criticisms in the literature.

The challenges to testing of Hotvedt, Hoffman and others were explored in Ravenwood, Ohio, a suburban school district where policy requires standardized IQ and achievement testing in grades three and six. They also require other kinds of tests from time to time, with an increasing number administered by the central pupil testing department. Beyond that, local option of the principal has left to teachers considerable discretion in testing practice. The district test scores of Ravenwood are well above the national average. Its middle to upper class children are generally regarded as winners, the kind of children the schools and testing system are designed to benefit.

All the quotations in this study are reported from interviews with three early childhood teachers conducted at Woodhill Elementary School in November 1975. Mrs. Bernstein teaches two half-day sessions of kindergarten, Mrs. King first grade and Mrs. Langley as combination class of first and second graders. All three teachers are experienced, dedicated professionals. Their classroom can be characterized as informal, semi-structured, and pluralistic--rich in opportunities to learn with students participating in a variety of activities simultaneously.

The interviews were opened with, "I'm interested in the uses teachers make of tests and test results. Testing means anything you define as testing or a way of getting information about children, from standardized tests to any kind of teacher assessment. Do you use tests?"

The ensuing interviews were taped for later analysis. The discussions focused primarily on the seven issues which emerged during the interviews: evaluation policy in an elite community; cooperative policy making for influencing change; harmful effects of testing; fairness; individuality and standardized tests; teacher assessment vs. standardized testing; and evaluation integrated with curriculum. Further analysis has synthesized these recurring themes into the three predominant issues: evaluation policy; harmful effects of testing; and individualized evaluation.

Evaluation policy: "Admission to Harvard and super gain scores"

As schools tend to reflect the communities they serve, testing policies at Woodhill School appear to reflect the values of the community's parents. The teachers think many parents attained Ravenwood status by improving on the economic and social status of their own parents. This status may in part be due to more education. Testing has such an impact on schooling that the teachers say the parents tested themselves upward into Ravenwood.

The teachers regard many Ravenwood parents as elitists with the highest expectations for their children, citing such unrealistic goals as admission to Harvard or "super gain scores." Parents' goals have created a climate in which demands for accountability exert pressures on the school administration. "We have a 200% PTA," noted Mrs. Bernstein. "It is possible we could have 500%. They are overly, super involved."

Demands for accountability and recently purchased programmed curriculum materials with tests embedded in them both contribute to escalating the amount of testing. This creates pressure on teachers and children. Mrs. King fears that accountability pressures and resulting tests or paper work may increase anxiety, teaching children to hate school. She feels the situation is compounded by the administration policies that "try to meet children's needs by being accountable to parents." Mrs. Bernstein is concerned that parents "worry and get uptight," and may misinterpret scores to mean their child is less bright than they would like, or that he is enrolled in the wrong school, with less competent peers. These feelings ultimately are transferred to the child.

The teachers would like to communicate more closely with administrators and parents on evaluation policy, the realities of its practice, and its effect on students. "The trouble is, the district makes these policy changes without asking the teachers, and it's hard to work together, even with the specialists."

Mrs. Bernstein holds that the "concept of primary education has changed. In the 1960's policies were freer. Teachers had more responsibility for developing curriculum and there were more real life experiences, more activities and trips as the basis for learning. There was little concern about academic skills and assessment for first grade readiness." Now, the teachers are not consulted with regard to the district's selection of required curriculum materials and kits or the tests associated with them although they have the responsibility to implement these curricula and administer the tests.

Teachers and parents may share similar hopes for children's development, but differ in their approaches. An example of working more closely with parents is Mrs. Langley's approach of enlisting parents as classroom volunteers to help them understand more about child development, learning and within reason what children can or cannot do.

The teachers resist evaluation policy that is based on standardized testing because they feel it is unfair and minimized individuality. They would like to avoid using tests, claiming that harmful effects on children outweigh the advantages to parents or teachers.

Harmful effects: Teaching for creativity and divergent thinking, measuring convergence

The testing situation is viewed by these teachers as "socially and emotionally damaging for children to experience." Mrs. Langley said, "There are so many variables in a test situation you can't give the same test to everyone. One may not have gotten started. Movement distracts some. Maybe a child didn't have breakfast. Maybe the test is going to tell him how he is incompetent or destroy his self-image and he knows that. You know tests test for the average, and the way they are scored, 50% have to be below the norm." Mrs. King describes a problem with new kindergarten tests that are used to predict performance and determine pupils' placement in specific programs in first grade. "The tests...include such difficult skills that only a very unusual child could get very far into them. That is hard on their self-esteem as they take the test."

Critical of tests themselves, the teachers object to the narrow range and selection of skills, variables and "not quite right answers" that are acceptable on test items. The teachers see the children in more complexity than tests can measure. For this reason, they are skeptical of tests' accuracy and usefulness. Children are developing and changing all the time, even if idiosyncratically. "We do not know anything for sure, even with pre-test, post-test procedures," said Mrs. King.

"We teach for creativity and divergent thinking when tests measure convergence. Tests freeze children in a score even though they go right on developing."

Labeling children with results of questionable test items is basically unfair and too controlling. The labels placed on them in the early grades bias subsequent teachers' expectations, the peer relationships, the parents, and the children's self-expectations. "The self-fulfilling prophecy works," Mrs. Langley is convinced. Eventually, test scores and labels track children educationally into classes from which it becomes increasingly difficult to extricate themselves. Early childhood teachers, they asserted, should be reluctant to begin this process of stigmatizing children.

"We want to give children every opportunity to learn and grow without putting labels on them. I try to give all those opportunities before I'm willing to call others to come in here and take the child apart," said Mrs. Bernstein. "I believe so much learning has to do with maturational experiences and growing. Yet, we are tracking children early with standardized tests."

The teachers object to sharing their power with testing specialists, they think that they abuse that power. Teachers feel it is an invasion of their territory, devastating a percentage of the children by lowering their self-esteem and leaving the teacher to pick up the pieces. "The real question," added Mrs. King, "is how and who uses the data from tests." Not only are teachers unable to obtain test results they wanted to use for diagnosis and remediation follow-up, but specialists with access to the data have also failed to make it work for children. Mrs. King holds that "at the end of the year testing specialists sent lists of children who did poorly on this test. They announced they would work individually with those children in perception. We never saw them." All teachers reported getting results back so late that students were already doing the skills the tests showed they could not do, especially in the area of perceptual skills.

"Testing doesn't help the child. It helps the teacher, but we get the results back so late," said Mrs. Bernstein. "The follow-up is really the teacher's responsibility." Results are used for political and social purposes in the schools and the community, but diagnostic uses are thwarted. Mrs. Langley argues testing "is not useful, is costly and very time consuming."

To alleviate the futility of gathering results later misused, they advocate an open communication arrangement with the testing office to establish that particular children's problems

have diminished. "That's one way to reduce the stigma," said Mrs. Langley. In general, the teachers say they have seen little evidence of constructive use of standardized testing to outweigh the harmful effects.

Individualized evaluation: "Did it work with this child?"

Alternative ways of assessing children in the context of the learning situation are being developed by the teacher as an integral part of their teaching. School can make a difference, they believe, if children are viewed as individuals. The teachers see evaluation as an ongoing, formative part of the classroom process. They emphasize the relationship between the quality of content in the learning activities and the quality of what individual children do to learn and produce. The basis of early childhood curriculum for these teachers seems to be the teaching of powerful concepts. To promote the development of a real understanding they offer children a variety of learning experiences for each concept. Their most frequent questions to themselves are, "Did it work with this child? What else could I do to make it more effective? Can he also do this alternative task?" The teachers say they try to cross-check a child's ability to grasp the concept by his performance in several alternative tasks.

What the teachers think is worth evaluating and their methodology are tied together. In attempting to individualize both instruction and evaluation, frequent observations and interviews with children guide the teacher's day to day planning. Assessing what and how well children are doing, at what skills they are successful or weak, indicates to the teacher what is appropriate for the child to try next. "There is so much variation in children and the way they learn. A teacher can accommodate a lot of different right answers from different children for any given concept, which a test can't do. You can also retest in a sense when you help them move to the next step," said Mrs. Langley.

Mrs. Langley described her method of watching the child's learning style, how self-directed the child is; how he makes choices, plans and decisions about "what he thinks the next step is. Maybe that's motivation. You can't test it, but it's here to observe," she said.

Mrs. King recalled, "I watch and observe and jot down notes on them--anecdotes, thoughts, comments--once a week or so, sometimes once a day. I look at social and emotional development, attention span, maturity, physical development, perception, academic progress, ability to make decisions, to analyze, discuss and to ask questions. We do a lot of sharing and discussion. It's interesting how accurately the kinds and variety

of questions, answers and dialogue I'm getting are indicators of where the child is. Giving direct academic help or social/emotional help is an enigma, because you can't really tell the direct results. The answers seem to lie in providing a huge variety of experiences, rather than in pegging the children."

The teachers prefer information gathered in the context of a more wholistic evaluation of the child's learning by someone who knows him and works closely. They try to consider their teaching and the environmental as well. They have developed their own kind of tests using categories of subject areas and abilities they find relevant to learning and teaching. They observe, interview and make judgments about achievements in particular situations. Using anecdotal records, they gather comparative data on each child on individual skills. Often they focus on such evaluative criteria as level of participation, level of conceptual complexity, ability to verbalize (spoken or written) or make drawings about learning experiences, and improvements in language and discourse.

Occasionally they use a checklist of various skills developed partly from the testing. Mrs. Bernstein acknowledged "that the perceptual testing improved my teaching, made me more aware and observant about perceptual elements in the child." The teachers gave examples of how they evaluate children's work, and illustrated making inferences about perception, motor sensory development, reading, math and interpersonal relationships.

For reading assessment, Mrs. Langley described her procedure. "I'd take the child to the library to pick out books he was interested in and go from there. I'd confer with him as he read, and work until I found any special problems." The relationship between language development and reading is regarded as so critical by these teachers that they stress experiential reading. The evaluation most pertinent is program and child specific. They evaluate vocabulary, syntax, and comprehension, using phonetic or sight approaches as appropriate to the individual. The teaching corollary is the emphasis on reading and using many books, and offering chances for discourse in the classroom. Mrs. Bernstein said, "Field trips give you dialogue, questions, direct experiences, and then all the real thinking. The sense of authorship from recalling and writing stories brings kvell. That means to be so pleased inside that you're smiling with your own accomplishments."

The math evaluation also consists of working individually with objects and activities in which math experiences are embedded. Their teaching stresses manipulative experiences prior to paper and pencil. Symbolization, geometric concepts, sets and operations are all taught and assessed with the same materials. Mrs. King believes understanding the concepts behind

numbers far more natural for kindergarten aged children than learning sounds and phonics. "If a child has a pretty good idea about numbers including the operations, I can be pretty sure he or she will probably function quite well in other areas too, especially reading."

Children's paintings and drawings are analyzed regularly for content. "They represent how the child feels about the world and himself. I get a sense of emotional maturity, of language development, of motivation, and of motor-sensory and perceptual development," noted Mrs. Langley, "but especially of aesthetic awareness."

Motor perceptual abilities are assessed by direct observation of physical abilities in hopping, skipping and jumping. Such aptitudes as spatial relationships are judged by behaviors like bumping into objects, clumsiness, or how a child sits in the group. Some children are disoriented in the class or school. Mrs. Bernstein gives practice to such children by sending them on missions to other places in school.

The teachers argue that it is important to evaluate children's interpersonal relationships--the ability to trust, respect each other and communicate. Those with interpersonal problems often develop learning problems later. For certain children with problems, the advantages of a wholistic evaluation approach are particularly evident, and point up the inadequacies of cognitive, test-based-assessments. Such children are sometimes strong academically, but need to be referred for special psychological services.

The results of this ongoing, informal teacher evaluation are reported to parents in ways that emphasize children's individuality. They keep a portfolio of anecdotal notes and observations, examples of the child's written work, and before and after drawings from beginning to end of the year. They also present a "progress report" at the conference about performance, abilities and idiosyncracies. "It is a summary of what I know of the child in a written case study," reported Mrs. Langley. Teachers prefer analytic description to normative report cards, they said. Referring to the "myth of test norms" they regard report cards as competitive and unfair to the children.

The teachers are alert to how they use elements in the curriculum as indicators for evaluation, but they see this as very different than teaching to tests. Their indicators are varied and situational, the assessment is conducted in a naturalistic setting, and it is supportive rather than threatening. "It is no different than what we do every day, so that all children have the equal chance to do their best."

The teachers trust their own assessment of children in the context of the classroom as valid, more so than tests. In contrast to the teachers' individualized evaluation, standardized tests seem to them to deny certain individual characteristics and circumstances. "It seems to me we've seen over and over again that teachers can evaluate children as well or better than tests, and since they do it in a supportive way it is not so damaging." The judgments they make are related to their teaching, and that "makes it more fair to the child because it improves the quality of what happens for each child in the classroom," claimed Mrs. King.

Parents in Ravenwood expect their children are college bound. If they want teaching for creativity and divergent thinking, if they want direct field experiences that develop dialogue and question-asking, that produce the thinking children can write and read about, it takes a certain style of teaching. That style requires a harmonious and appropriate style of evaluation that is embedded in the teaching, the teachers asserted.

They see individuality and growth partly as a function of a child's response to many diverse learning experiences intended to extend knowledge, creativity, and aesthetic sensitivity and critical thinking. Because they cannot predict all the learning that will take place, they try to provide diverse activities, "If we were to narrow teaching to basic skills that are test items, we would never know what talent we missed, what characteristics could have developed, what appreciations, what relationships," concluded Mrs. King.

REFERENCES

Anderson, G.J. and Walberg, H.J. "Learning environments" in Walberg, H.J. (ed.) Evaluating educational performance. Berkeley: McCutchen, 1974.

Barclay, J.R. "Need assessment," in Walberg, H.J. (ed.) Evaluating educational performance. Berkeley: McCutchen, 1974.

Bussis, A.M. and Chittendon, E.A. Analysis of an approach to open education. Princeton: ETS, August 1970.

Bussis, A.M., Chittendon, E.A. and Amarel, M. "Alternative ways in educational evaluation," in Testing and evaluation: new views. Washington, D.C.: ACEI, 1975.

Bussis, A.M., Chittendon, E.A. and Amarel, M. Beyond surface curriculum. Boulder: Westview Press, 1976.

Carini, Patricia. "The Prospect School: taking account of process," in Testing and evaluation: new views. Washington, D.C.: ACEI, 1975.

Duckworth, E.R. A field and planning study for the appraisal of reading in open classrooms. Princeton: Early Education Group, ETS, 1973.

Eisner, E.W. "Instructional and expressive educational objectives: their formulation and use in curriculum," in AERA Monograph Series No. 3. Chicago: Rand McNally, 1969.

Eisner, E.W. The perceptive eye. Mimeo paper presented to Division B., AERA, 1974.

Elliott, J. and Adelman, C. "Reflecting where the action is: the design of the Ford Teaching Project," Education for teaching. Cambridge Institute of Education, 1973.

Elliott, J. Evaluating the Progress in Learning Science dissemination. mimeo, Cambridge Institute of Education, 1976.

Gallagher, J.L. "A topic classification system for classroom observation," in AERA Monograph Series No. 6. Chicago: Rand McNally, 1970a.

Gallagher, J.L. "Three studies of the classroom," in AERA Monograph Series No. 6. Chicago: Rand McNally, 1970b.

Glass, G.V. "Teacher effectiveness." in Walberg, H.J. (ed.) Evaluating educational performance. Berkeley: McCutchen, 1974.

Hastings, J.T., Runkel, P.J., Damrin, D.E., Kanes, R.B., Larson, G.L., The use of test results. DHEW Cooperative Research Project No. 509. Urbana, IL.: University of Illinois, Bureau of Educational Research, 1960.

Hoffman, Banesh, in Houts, P.L. "A conversation with Banesh Hoffman," National Elementary Principal, 54, (6), August 1975.

Hotvedt, M.O. A case study of standardized test use in the public schools. Unpublished doctoral dissertation, University of Illinois, Urbana, 1974.

House, E.R. "Technology and evaluation." Educational Technology, November 1973, 13, (11), 20-26.

Karier, C.J. "Testing for control and order in the corporate state." in Karier, C.J. et al., Roots of crisis: American education in the twentieth century. Chicago: Rand McNally, 1973.

Kemmis, S. Telling it like it is: the problem of making a portrayal of an educational program. mimeo, CIRCE, University of Illinois, Urbana, 1974.

Perrone, Vito and Strandberg, W. "A perspective on accountability." Teacher's College Record. February 1972.

Perrone, Vito. "Alternatives to Standardized Testing," National Elementary Principal, 54, (6), August 1975.

Scheyer, P.T. and Stake, R.E. "A program self evaluation portfolio," Studies in Educational Evaluation, 2, (1), Spring 1976.

Stake, R.E. "The countenance of educational evaluation." Teacher's College Record, 68, April 1967.

Stake, R.E. "Testing in the evaluation of curriculum development." Review of Educational Research, 38, (1), February 1968.

Stake, R.E. Estimates of demonstrated validity of various kinds of tests used for various educational purposes. unpublished mimeo. CIRCE, University of Illinois Urbana, 1975.

Stake, R.E. Program evaluation, particularly responsive evaluation. Occasional Paper No. 5. Kalamazoo: The Evaluation Center, Western Michigan University, November 1975.

Wolf, R.L. Making education accountable to the learner: a framework for evaluation. mimeo, CIRCE, University of Illinois, Urbana, 1971.

Zacharias, J.R. "The trouble with IQ tests." National Elementary Principal, 54, (4), March 1975.

CONVERSATIONS WITH MYSELF

OR

WHAT'S A NICE GIRL LIKE YOU DOING IN EVALUATION?

Carol Hodges Wardrop
Elementary Education

Preface

Once upon a time a long time ago (August 1973) a "rather trusting, naive, unsophisticated, gullible and unworldly" girl entered the lion's den (otherwise known as the real world of evaluation!). She was never seen or heard from again. Some folks say they've seen a "worldly, sophisticated, scheming shrewd and experienced" lady who looks vaguely familiar near the entrance to that same den. But that couldn't be her...could it?

Hmm...I think I see her now, I wonder.

"Excuse me, 'worldly, sophisticated, scheming shrewd and experienced' lady. You wouldn't happen to be that 'rather trusting, naive, unsophisticated, gullible and unworldly' girl who entered the den two years ago, would you?

You would!! What was a nice girl like you doing in evaluation???"

"Sometimes I wonder about that myself. The naive girl that I once was certainly did not expect evaluation to be such a socio-political-economic-administrative-legal-moral conglomeration that I now know it to be. Have you ever heard of an evaluator named Dan Stufflebeam? He wrote a paper on meta-evaluation in which he discusses six classes of problems that he believes may jeopardize an evaluation study if they are not thought through before the study begins. These classes are conceptual, socio-political, contractual/legal, technical, administrative and moral/ethical. If you really want to know what happens to a nice girl in evaluation, I'll have to discuss a few of those problems.

In order to understand my story, I'll have to take you back to 1972 when the American Education Foundation (AEF)* awarded contracts to two ambitious efforts to implement large-scale and long-term demonstrations of their capabilities as systems of computer-assisted instruction. One of

*Although this story is true, all names have been changed.

them, Information Computer Television (ICT), was being carried out by the RET Corporation in Southcity. The other, known as Logical Information Programming (LIP), was being developed by the Computer Laboratory of Big University. The Testing Service of America (TSA), located in Easternton, was awarded a contract by AEF at the same time to conduct an extensive and thorough evaluation of each of the two systems.

The evaluation of the ICT and LIP programs was designed to focus on three major components of the demonstration projects--their costs, performance and educational effectiveness. The classroom demonstrations and the evaluation of them were expected to begin in September 1973 and continue for approximately two-and-a-half years. The contract called upon TSA to send quarterly, annual and specified special reports to AEF on a regular schedule. In addition, RET Corporation and Big University were to receive copies of the reports. By May of 1973, TSA had drafted a document describing a plan for the evaluation. This document presented, among other things, the role that TSA was to play in the evaluation, goals and behavioral objectives, sampling plans, instrumentation, data analysis, a timetable and dissemination of the results. Conceptually, I believe it was very well thought out."

"Okay, okay. That's all well and good. But where are you in this picture 'trusting, naive, unsophisticated, gullible and unworldly' girl? What was happening to you? What part did you play? What was happening out in the real world of evaluation?

"Oh yes, I did promise to tell you what happened, didn't I? It's just that it's so much easier to talk about plans than it is to make them work. Extraneous things like teachers, students and machines that don't work keep getting in the way. I was only involved with one component of the evaluation, so my conversation with you will have to focus on the evaluation of the elementary school math and reading portion of the LIP Evaluation. I really know very little about the other parts of the LIP evaluation or anything about the ICT portion.

Because TSA is located in Easternton, the evaluators there decided to hire someone here in Middletown to act as a local evaluation coordinator. That became my job. My initial task was to talk with the LIP staff who were working on the elementary programs in order to ascertain the state of the curriculum. I was also to discuss with them their plans for involving local schools and teachers in the field trials, which were expected to begin in a month. I also became acquainted with the computer language and the LIP computer facilities. During my first week on the job, it

became apparent that anyone associated with the evaluation and "THAT PLACE" (TSA) were suspect. I had no place to work, no desk, no equipment and no supplies."

"Oh, oh! You ran smack into one of the administrative problems Stufflebeam talks about didn't you? TSA had made no prior provision for office space, equipment or materials? What was available? Answers to questions like that can really effect the ease with which evaluations can be carried out. Is that when you started to change into that 'worldly, sophisticated, scheming, shrewd and experienced' lady?"

"No, not yet. I was still rather naive and believed in things like the TSA 'Fairy Godmother.' Anyway, I did finally get a desk, even though I had to move three times that first year. (A year later when another local evaluator joined me, she only had to move twice!) Unfortunately we never did get a typewriter or secretarial help. Much of our time is wasted Xeroxing reports which we send to TSA to be typed. I have become rather fond of the Xerox machine though...when you are labeled as an evaluator, you sometimes have to find your friends in strange places!!! You are right though. Administrative problems such as location and facilities should be discussed and taken care of before the evaluation begins.

Anyway that was just a small problem. After working for a week or so at the Computer Laboratory, it became obvious to me that the field demonstrations would not begin as scheduled. The LIP personnel had not even recruited teachers yet. Plans for the implementation had changed frequently since February 1972 and courseware developments had slipped seriously. These uncertainties complicated the evaluation plans. Without more complete information on the implementation, including the distribution and placement of terminals, content outlines of all instructional materials, and identification of teachers and pupils who would use the system, TSA's efforts were severely hampered."

"I guess you are right. How can you evaluate a program if there isn't anything there to evaluate. What did the contract say about protecting the evaluation if there were delays in the product? Doesn't Stufflebeam say all parties involved in an evaluation require protection and assurance? What happened? Don't stop now, its just getting interesting."

"Interesting for you maybe, but a headache for us. Plans for the evaluation had to be changed. It seems that although LIP had been granted a year's delay for their field trials, AEF was still expecting TSA to have all of their data to be gathered by December 1975. What had originally been a formative and summative evaluation was turning into a formative one only. Something should have been written into the contract to protect TSA in the case of delays, but it hadn't so we tried to evaluate what was ready.

Finally in early 1974 a few teachers in two schools began to use the computer material on a very limited basis. This was termed a "mini-field trial" by LIP. My job became one of observer. I sat in the classrooms involved and took copious notes on everything that was happening. These narratives were sent to TSA and a classroom observation instrument was later drafted from them. I tested the instrument, suggested changes and sent in more narratives. Things proceeded like that for the rest of the school year.

In mid 1974 a revised version of the evaluation design was presented, which was conceptually and technically well thought out. An important influence on the plan was a concern for the plurality of values exhibited by the various audiences. It was recognized that no plan can be equally responsive to all audiences. Nevertheless, an attempt was made to provide those persons, groups and institutions having an interest in the outcomes of LIP with information relevant and appropriate to their varied interests. To be responsive to the goals and characteristics of the LIP elementary school project itself required considerable custom tailoring of procedures and instruments. On the other hand to be responsive to the needs of those outside the project who were interested in computer-based instruction required traditional measures and designs as well.

Demonstrations of the elementary school component were to be conducted in schools in Middletown. Which schools would participate was a function of which teachers in the school districts volunteered to participate and the availability of terminals and peripheral equipment needed for their operation. Thirteen reading and nine math teachers were involved during the 1974-75 school year. (By September 1975 ten new reading teachers and three new math teachers had joined the ranks of the volunteers.)

The participants in the evaluation were not limited to those who had volunteered to use LIP in their classrooms. It was expected that comparison classes would be provided by

the school districts as part of their obligation to the demonstrations."

"I notice that you used the phrase 'expected that comparison classes would be provided.' Does that word 'expected' mean anything? Sounds like another contractual problem... definition of roles?"

"That's right. Although TSA expected volunteer comparison classes it was never in any legal document. During the 1974-1975 school year some comparison classes were obtained, but usually because a school administrator had 'volunteered' a teacher (often without that teacher's prior knowledge)."

"But...didn't you realize that unless people who will need to support the evaluation are involved early and meaningfully, they are not likely to support it?"

"We knew it. There was just nothing that we local people could do, because TSA handled those negotiations. The comparison teachers resented every minute they had to give up to our testing. This school year the same thing happened. TSA did not contact school officials until the last minute and some teachers were not notified that they would be participating in the evaluation as comparison classrooms until the week school started."

"I know AEF wanted the evaluation to see if LIP had any effect on the achievement and attitude in the classroom. How did TSA plan to go about that?"

"In the 1974 evaluation design, fourteen key issues were explicated. TSA decided to collect data relevant to those issues. The sources and modes of data acquisition used were structured in terms of the principal targets of the evaluation: the child, the classroom, the context of the demonstration, and the courseware. The measures related to the child were primarily those assessing achievement and attitude. Student work samples were also to be included, as were other measurement techniques such as parent questionnaires and school cumulative record data. Unfortunately, when it became apparent that the 1974-75 school year would only be a trial run the work sample and questionnaires had to be dropped. Resources had to be spread over a longer period of time than originally been planned.

In addition to collecting information on the impact of the system, it was also deemed our job to monitor the demonstration in progress. Information concerning strategies of implementation was considered essential to enable us to identify classrooms in which the program received a fair trial, and to describe those forms of implementation most conducive to various outcomes. A series of classroom observations were required to monitor this."

"Observations...I thought you'd never get to them. That's basically what you were hired to do, isn't it? There are a jillion observation schedules around, did you try any of them?"

"Yes the other local evaluator and I were originally hired as observers. According to the original observation plan we would receive an observation instrument, be trained in its use and go out and observe!! That isn't quite how it turned out though. As you remember, I mentioned earlier that in the Spring of 1974 I sent TSA many narratives which they used to fashion an observation scheme. In the Fall of 1974 the other local evaluator and I received copies of that instrument, which we used, revised, used and revised again. We trained ourselves on the job. In addition to working on the instrument for observing the whole classroom, we also fashioned one for observing the child at the terminal. After almost a year's work, we finally produced two instruments with which we were partially satisfied.

The hardest part about observing though is writing it up. It seems that no instrument says everything, so after one finishes all the checks or circles or whatever one is using, there is always something important left over to write a narrative about."

"Let's see, didn't you mention something about on-line collection of data? That sounds like a good source of information. I've heard that computers can be programmed to collect all kinds of things. What are you going to do about that? I don't think you've mentioned anything about the courseware development data either. Didn't the evaluation have anything to do with that too?"

"Yes the evaluation was to monitor both the on-line data collection and the courseware development data. However, neither of those areas was under my jurisdiction."

Recent federal legislation made it very difficult for us to obtain much of the material which we wanted from school records. Certain data in school files is now considered confidential except to the student or the student's parents. This has meant that though we may obtain a general picture of the socioeconomic status of the children who are in certain classrooms we may not ask about each child individually, unless we get written permission from the parents and this type of permission gets harder and harder to obtain.

LIP was being considered by the evaluation as constituting an additional material resource to the classroom that was expected to have an impact on how other resources in the room would be perceived and used. The nature and interplay of classroom resources was to be monitored through several procedures, each directed at selected aspects of the classroom and using a methodology appropriate to its main purpose. There were to be six major modes of data collection involved in this stage: work diaries and program summaries, teacher logs, teacher interviews and questionnaires, classroom observations, individual pupil observations at the terminal and on-line data collection."

"No wonder you became a shrewd, worldly, scheming lady!! Anyone who could convince a classroom teacher to help with those things has to be scheming."

"That is true. Actually though, we did not collect all of that data. It was just in the evaluation plan. When resources were cut back due to LIP's delay, the work diary, program summary and teacher questionnaires were lopped off the plan. We did attempt the rest of the techniques. The teacher log was one which was worthwhile in some cases. The teachers were asked to record changes in the patterns of their activities and preoccupations, shifts in their priorities and changes in academic or social process attributable to LIP."

"Those teacher logs sound like a good idea, how did they work out? I'll bet some teachers did a super job and others hardly wrote anything. Did you get any information that you might not otherwise have received? Was it worthwhile?"

"Yes, we did find some interesting comments. Our only problem is motivating teachers to write in the logs on a regular basis. We don't have time to check up on them so we just have to hope teachers are keeping up with them. The other technique which provided us with interesting information is the teacher interview. We found teachers to be very open with us and eager to provide us with their views on LIP and how it effected their classroom.

"Well it sounds like you had enough to do anyways. Let's get back to that real world of evaluation again. If I remember correctly, you had been working at this job for a year now and the real field demonstrations had not begun yet. That was supposed to happen soon though wasn't it? Is this when you become a full-fledged evaluator?

"I'm not sure if you'd call it full-fledged or not, but that was when some of the naiveness, truthfulness and un-worldliness began to rub away. Even a year after the original starting data the Computer Laboratory people still were not ready. They did have some classrooms now and teachers, but the LIP materials were continuing to be developed and modified. The terminals were not reliable. Things were in a constant state of flux.

At least we did have some classrooms to observe and with this realization it became apparent that I could not cover 21 classrooms by myself. This was when TSA hired the other local evaluator whom I've mentioned before. We tried to observe and begin the pre-testing at the same time. Obviously we were spreading ourselves too thin, so we hired a local person to do the testing for us."

"I'm way ahead of you this time. Another of Stufflebeam's problems appeared. I can even guess...it is staffing. He says it is important to decide who will have overall responsibility for the work before an evaluation begins. He also says that questions like...what other roles are to be manned (personnel)? ...what recruitment of personnel will be done? ...should be answered ahead of time. Am I right?"

"Unfortunately you are right. From the onset of the evaluation effort, no one at TSA had overall responsibility for the evaluation. Everyone who was involved from Eastern-ton had this evaluation as only one fraction of his work. Information is promised to the three of us who work locally but it never comes. The three of us who were originally hired as observers and testers have become administrators and writers as well. Don't get me wrong, it's not responsibilities that bother us; we rather enjoy the challenges. What does bother us is that we have had no prior training for the work we are doing."

"Another problem, huh? Stufflebeam says that if the various persons are to perform their roles effectively, they often need special training."

"You are right. Training is a key area in evaluation work. Much valuable time has been lost because training became a series of approximations until each of us finally learned the skills. Even more important was the complete lack of communication between the Easterton TSA group and the three of us. We need feedback not only for our own information but also to appear credible to others.

Enough of that...another problem soon popped up. This one involved the scheduling of data collection activities. No one at TSA had asked questions like: Is this schedule reasonable? Will the instruments be ready? Are there conflicts between the data-gathering and other events at the school? How long does it take to administer the tests? How many children can take it at once?

Just as important and equally unthought of were questions like: Does the budget reflect the evaluation design? How can it be changed if delays occur? What if new staff have to be hired? You can see the way management and economical problems play a part in the inefficiency of a project."

"I never realized that evaluating was so complicated. Where are you now? If I have been following you, it seems like you are in your third year on this job? Isn't everything supposed to be wrapping up? Have you had an extension? Are you still collecting a lot of data? Did you realize how much time it would take? What do the teachers think about having someone in their classroom testing all the time?"

"As you've already guessed it took much longer to administer tests than anyone could have estimated. In most classes testing conditions were bad. Unless evaluators consider and respond to questions about data gathering prior to its actual implementation, there are bound to be problems. It is hard to be unworldly and gullible when the world is telephoning you every night complaining about the tests, testers and conditions. A lot of shrewdness began to appear about now; I had to answer the complaints somehow!"

"Well, what have you found out about LIP? Are there any results?"

"Results? I wish I had a lot to tell you, but the truth is that I really don't know. The analyses performed on last year's data collection is done, but we have not received any of the results. We have received an extension on this year's collection and will be able to collect data through May. However, with this extension has come two changes in plans since August. We are never sure when we

are going to get a call from TSA saying that there has been another change and that we should scrap our schedule and begin anew."

"It is obvious now why you changed from that 'rather trusting, naive, unsophisticated, gullible and unworldly' girl to the 'worldly, scheming, sophisticated, shrewd and experienced' lady that you are now. The thing I don't understand is why do you stay with it? Regardless of all the changes which you claim have taken place in the last two years I still think you are a pretty nice person. Tell me what's a nice girl like you doing in evaluation?"

"The answer to that question isn't easy. Consider the evaluation plan I just told you about. It really had some fine things in it. The people at TSA who wrote it, were trying to find out about the effectiveness of LIP. Unfortunately, they did not always run things too well administratively. I guess I stay in this job because it is giving me a chance to learn things that I could never learn from a book. I stay in the field of evaluation because I believe that it is a very important one. Perhaps the experiences that I am having and hope to continue to have will allow me to help make things in the field better. Evaluation is changing. Look at all the new ideas which have been tossed around since Tyler's model. It is exciting...I want to be a part of it."

"POSTFACE"

"Well, that is the story. It seems to me that she really hasn't changed that much. She may be a little wiser, but still 'rather trusting, naive, unsophisticated, gullible, and unworldly.' Imagine thinking she can change evaluation."

